DSTO

Signal Processing Methods for Gearbox Fault Detection

Simon Rofe

DEPARTMENT OF DEFENCE
DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

Signal Processing Methods for Gearbox Fault Detection

Simon Rofe

Airframes and Engines Division
Aeronautical and Maritime Research Laboratory

DSTO-TR-0476

ABSTRACT

Methods of accounting for load variation in vibration signals from helicopter transmission systems are presented. These methods are based on autoregreassive moving-average (ARMA) models, and several ARMA parameter estimation schemes are presented. Simulations of load variation are carried out, and a prediction error filter, based on the ARMA models, is used to generate a residual signal. Fault indices extracted from the residential signal are used to indicate the presence or absence of a fault. The results of the simulations suggest that this method of fault detection is able to detect both general and local fault conditions.

DEPARTMENT OF DEFENCE

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

Published by

DSTO Aeronautical and Maritime Research Laboratory
PO Box 4331
Melbourne Victoria 3001

Telephone: (03) 9626 7000
Fax: (03) 9626 7999

AR-010-104
February 1997

APPROVED FOR PUBLIC RELEASE

Signal Processing Methods for Gearbox Fault Detection

Executive Summary

Current vibration-based techniques used for gearbox fault detection rely on analysis of signals captured under fixed load and fixed speed conditions. While helicopter gearboxes operate nominally at fixed speed, they do not operate under fixed load. If such fault detection techniques are to be implemented as part of self-contained on-board systems, then adaption to, or accounting for, load variation is a necessary requirement. This report details the signal processing methods which are able to account for load variation.

Several methods of linear modelling are introduced. These techniques enable the vibration signals recorded from helicopter transmission systems to be modelled. The aim is to identify a suitable model of the gearbox vibration signals under normal operating conditions, which is able to be adjusted to meet variations of the load imposed on the gearbox. Once a suitable model is established, the vibration signal recorded from the gearbox is compared with the signal produced by the model. Various signal characteristics (referred to as fault indices) are derived from this comparison as a means of deciding if a fault is present or not.

The results of several simulations show that this type of linear modelling provides a useful means of both accounting for load variation, and detecting the presence of faults.

Authors

Simon Rofe
Airframes and Engines Division

Simon Rofe received the BEng. (hons) degree in Electrical and Electronic Engineering from the University of Southern Queensland in 1991, and the PhD. Degree from the University of Queensland in 1996. He commenced work with the Airframes and Engines Division of the Aeronautical and Maritime Research Laboratory in 1995. He is involved in the area of gearbox fault diagnosis using vibration analysis.

Contents

Figures

THIS PAGE IS INTENTIONALLY BLANK

1 Introduction

Current vibration-based techniques used for gearbox fault detection rely on analysis of signals captured under fixed load and fixed speed conditions. While helicopter gearboxes operate nominally at fixed speed, they do not operate under fixed load. If such fault detection techniques are to be implemented as part of a self-contained on-board system, then adaption to, or accounting for, load variation is a necessary requirement.

Initial investigations of gearbox vibration signals indicate that load variation acts to change the signal characteristics in a non-linear manner. This implies that spectral and statistical properties of the vibration signal vary non-linearly with load. Consequently, adaption to load variation cannot be achieved by simply normalising the signal with respect to the load, or some signal feature which is proportional to load.

This report details signal processing methods which are able to account for load variation.

One of the difficulties in attempting to account for load variation in the vibration signal is that it is not possible to use a strictly adaptive method. Such methods usually derive an error signal from the actual vibration signal, and a mathematical model used to predict the vibration signal. This error is then used to adjust the model parameters in order to reduce the error power. A problem arises when a fault occurs in the system, in which case, the model will be adjusted not only to account for the variation in load, but also to any change in system dynamics due to the fault. The aim, then, is to account for load variations while maintaining the ability to detect faults.

2 Signal Model

The signal processing methods used herein essentially involve modelling the signal of interest, and adjusting the model parameters appropriately, to account for varying load conditions. The underlying model may be either (1) linear or (2) non-linear. While the theory of linear models is well established, non-linear models have found prominence in recent years [7, 9, 10, 11, 15, 24, 25].

There are many types of non-linear models including non-linear networks, bilinear models and Volterra series. Due to their complexity, and the computational burden, non-linear models should only be used if absolutely necessary. Conversely, linear models are much easier to implement, and require less computational effort. In this report only linear models will be considered, however, the techniques are readily extended to non-linear models.

One of the most general linear models is the autoregressive moving average (ARMA) model. Such models have been used in fields as diverse as economic forecasting, seismology, speech processing, array processing and adaptive filtering [13, 14, 16, 19, 20, 21, 22]. ARMA models are also widely used in control theory.

2.1 General Linear Signal Model

The signal under investigation is recorded from an accelerometer mounted externally on the gearbox. Since the measured vibration signal is stochastic (as opposed to deterministic), it is convenient to model it as a white noise process passed through a linear filter with a rational spectral density function as shown in figure 1. Here $\{x[n] : n = 1, 2, \ldots, N\}$

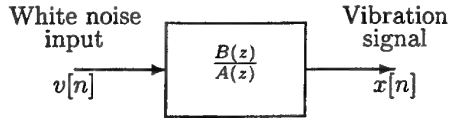

Figure 1: Vibration signal model.

is the recorded vibration signal sequence and $\{v[n] : n = 1, 2, \ldots, N\}$ is a Gaussian distributed random sequence with zero mean and variance σ_v^2 (often it is assumed $\sigma_v^2 = 1$ for convenience). The polynomials $A(z)$ and $B(z)$ represent the poles and zeros. respectively, of the linear system. If

$$A(z) = 1 \tag{1}$$

the model consists entirely of zeros and is known as a moving average (MA) process. Alternatively, if

$$B(z) = 1 \tag{2}$$

the model consists entirely of poles, and is known as an autoregressive (AR) process. Both $A(z)$ and $B(z)$ are assumed to be polynomials in z (the forward shift operator) of finite order, and share no common factors.

It is generally accepted [4, 13, 14, 16] that AR models are suited to processes which exhibit *peaks* in the power spectral density function, and MA models are suited to processes which exhibit *notches* in the power spectral density function. Although the AR and MA models (individually) are able to represent a broad range of signals over an arbitrary large frequency range, the ARMA model is able to achieve this with a smaller number of parameters.

It is important to note a fundamental difference between signal processing and control applications; viz., the properties of the signal $v[n]$. In general signal processing applications, access to $v[n]$ is unavailable, and it is assumed to consist of a Gaussian distributed process as described above. In control applications, however, $v[n]$ is often a deterministic function, of which access is generally available. Access to $v[n]$ makes the identification of the model $B(z)/A(z)$ simpler than if it is not available. Additionally, for Gaussian $v[n]$, $B(z)/A(z)$ is assumed to be a stable minimum-phase system (poles and zeros are inside the unit circle).

2.2 Fault Signal Model

The linear signal model above describes the process under normal operation (free from any fault). The fault may be incorporated by adding a signal which describes the action

of the fault,

$$x[n] = y[n] + \alpha\, z[n] \tag{3}$$

where $y[n]$ is the signal recorded during normal conditions, $z[n]$ is the signal due to a fault, and $\alpha = 0$ if normal operation or $\alpha = 1$ if operation under fault conditions. Figure 2 shows this graphically. The sequence $\{w[n] : n = 1, 2, \ldots, N\}$ is Gaussian distributed with zero

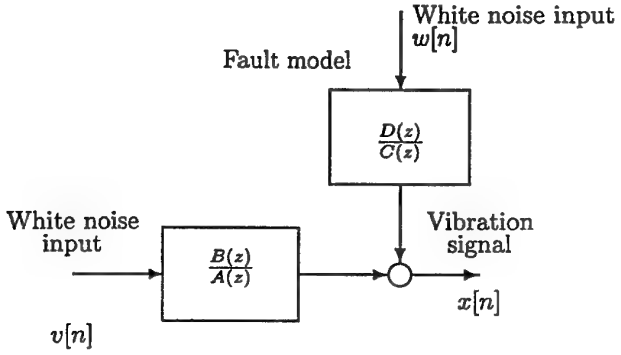

Figure 2: Fault signal model.

mean and variance σ_w^2. Additionally, $\{v[n]\}$ and $\{w[n]\}$ are assumed to be independent. The polynomials $C(z)$ and $D(z)$ describe the spectral characteristics of the fault signal sequence.

3 Accounting for Load Variation

The simplest method of accounting for load variation is to adjust the model parameters with respect to the load. This is illustrated in figure 3.

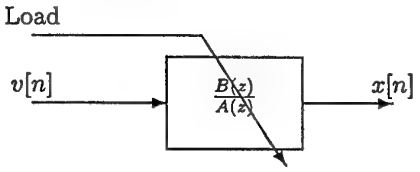

Figure 3: Adaptive parameter scheme.

This requires the model parameters to be identified as a function of the load. The model may then be written as

$$H(\zeta, z) = \frac{B(\zeta, z)}{A(\zeta, z)} \tag{4}$$

where the variable ζ is introduced to represent the dependence on load.

This scheme requires either the parameter variation to be determined as a function of load, or more simply, parameters to be identified for a series of different loads $\zeta[i] = L_{min}, \ldots, L_{max}$ where L_{min} and L_{max} are the minimum and maximum loads of interest.

It will be necessary to determine these parameters from vibration signals which are known to be fault free.

4 Parameter Estimation

One of the important requirements in using the modelling techniques described above is estimating the model parameters. These model parameters must be estimated from the gearbox vibration signal. There are several techniques for ARMA model estimation, and in the following four such methods are described.

It is well established [4, 13, 14, 16] that the estimation of the AR parameters involves a set of linear equations, whereas the estimation of the MA parameters is a non-linear problem. The latter of these is the more difficult to solve.

The first three methods presented below differ in the way they estimate the MA parameters of the model. They are also static estimators as they operate on a single data block. The fourth technique uses a recursive least squares algorithm to estimate the AR and MA parameters concurrently. It is also an adaptive algorithm since it updates the parameters continuously.

4.1 Kay and Marple Methods

The books by Kay and Marple [13, 16] present several methods of ARMA parameter estimation. Although the general motivation is spectral estimation, these methods are also applicable to system modelling and linear prediction. Two methods are presented in which the AR and MA parameters of the ARMA model are estimated separately. These are referred to as the modified Yule-Walker equations and the least squares modified Yule-Walker equations.

The general form of a rational polynomial ARMA model is

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + \sum_{k=1}^{q} b[k]v[n-k], \tag{5}$$

where $v[n]$ is a Gaussian distributed white noise process, and $a[k] : k = 1, 2, \ldots, p$ are the autoregressive parameters and $b[k] : k = 1, 2, \ldots, q$ are the moving average parameters for and ARMA(p, q) process. Multiplying equation (5) by $x[n-k]$ and taking the expected value gives

$$R_{xx}[k] = -\sum_{l=1}^{p} a[l]R_{xx}[k-1] + \sum_{l=1}^{q} b[l]R_{xv}[k-l], \tag{6}$$

where $R_{xx}[k]$ is the autocorrelation of the output sequence $x[n]$ and

$$R_{xv}[k] = \mathrm{E}\left\{v[n]x[n-k]\right\} = \mathrm{E}\left\{x[n]v[n+k]\right\}. \tag{7}$$

Now $R_{xv}[k] = 0$ for $k > 0$, since $x[n] = \sum_{l=-\infty}^{n} h[n-l]v[l]$ is the output of a causal filter and clearly depends on $\{v[n], v[n-1], \ldots\}$ which is uncorrelated with $v[n+k]$ for $k > 0$. Thus

$$R_{xx}[k] = \begin{cases} -\sum_{l=1}^{p} a[l]R_{xx}[k-l] + \sum_{l=1}^{q} b[l]R_{xv}[k-l] & \text{for } k = 0, 1, \ldots, q \\ -\sum_{l=1}^{p} a[l]R_{xx}[k-l] & \text{for } k \geq q+1 \end{cases} \tag{8}$$

The above relationship shows there is a non-linear relationship between the autocorrelation function and the ARMA parameters. This is due to the $\sum_{l=1}^{q} b[l]R_{xv}[k-l]$ term. If only the equations for $k > q$ are used, then this term may be omitted. This may then be written as the set of linear equations

$$\begin{bmatrix} R_{xx}[q] & R_{xx}[q-1] & \ldots & R_{xx}[q-p+1] \\ R_{xx}[q+1] & R_{xx}[q] & \ldots & R_{xx}[q-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}[q+p-1] & R_{xx}[q+p-2] & \ldots & R_{xx}[q] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = - \begin{bmatrix} R_{xx}[q+1] \\ R_{xx}[q+2] \\ \vdots \\ R_{xx}[q+p] \end{bmatrix}. \tag{9}$$

These are known as the modified Yule-Walker equations. These equations involve the actual autocorrelation function which is not generally known. A reasonable approach is to replace the theoretical autocorrelation function by an either the biased estimate

$$\hat{R}_{xx}[k] = \frac{1}{N} \sum_{i=1}^{N-k} x[n]x[n-i], \tag{10}$$

or the unbiased estimate

$$\hat{R}_{xx}[k] = \frac{1}{N-k} \sum_{i=1}^{N-k} x[n]x[n-i]. \tag{11}$$

4.1.1 Modified Yule-Walker Equations

The modified Yule-Walker equations may be solved by using an extension of the Levinson [13, 16] algorithm to give estimates of the AR parameters $\{a[1], a[2], \ldots, a[p]\}$. The recursive algorithm is initialised by setting

$$\begin{aligned} a_1[1] &= \frac{-R_{xx}[q+1]}{R_{xx}[q]}, \\ b_1[1] &= \frac{-R_{xx}[q-1]}{R_{xx}[q]}, \\ \rho_1 &= (1 - a_1[1]b_1[1])\, R_{xx}[q], \end{aligned} \tag{12}$$

Then recursion for $k = 2, 3, \ldots, p$,

$$\begin{aligned} a_k[k] &= \frac{-R_{xx}[q+k] + \sum_{l=1}^{k-1} a_{k-1}[l]R_{xx}[q+k-l]}{\rho_{k-1}}. \\ a_k[i] &= a_{k-1}[i] + a_k[k]b_{k-1}[k-i] \quad i = 1, 2, \ldots, k-1. \end{aligned} \tag{13}$$

If $k = p$, the recursion is stopped, otherwise

$$\begin{aligned} b_k[k] &= \frac{-R_{xx}[q-k] + \sum_{l=1}^{k-1} b_{k-1}[l] R_{xx}[q-k-l]}{\rho_{k-1}}, \\ a_k[i] &= b_{k-1}[i] + b_k[k] a_{k-1}[k-i] \quad i = 1, 2, \ldots, k-1, \\ \rho_k &= (1 - a_k[k] b_k[k])\, \rho_{k-1}. \end{aligned} \tag{14}$$

The solution is then $a[k] = a_p[k] : k = 1, 2, \ldots, p$, and the error variance $\sigma^2 = \rho_p$.

Following the estimation of the AR parameters, the data are filtered with the estimate of $A(z)$,

$$\hat{A}(z) = \hat{a}[1]z^{-1} + \hat{a}[2]z^{-2} + \ldots + \hat{a}[p]z^{-p}, \tag{15}$$

giving,

$$y[n] = -\sum_{k=1}^{p} a[k]x[n-k] \tag{16}$$

Durbin's method [13, 16] is applied to this filtered data to estimate the MA parameters $\{b[1], b[2], \ldots, b[q]\}$ and the white noise variance σ^2.

Durbin's method consists of the following two steps,

1. Using the data, $\{y[1], y[2], \ldots, y[N]\}$, fit a large order AR model using the autocorrelation method (based on the Levinson algorithm). For an AR model of order L, where $q \ll L \ll N$, the white noise variance estimator is given by

$$\hat{\sigma}^2 = \hat{R}_{xx}[0] + \sum_{k=1}^{L} \hat{a}[k] \hat{R}_{xx}[k]. \tag{17}$$

2. Using the AR parameter estimates obtained in step 1 as the data $\{\hat{a}[1], \hat{a}[2], \ldots, \hat{a}[L]\}$, use the autocorrelation method with an order q to find $\{\hat{b}[1], \hat{b}[2], \ldots, \hat{b}[q]\}$.

4.1.2 Least Squares Modified Yule-Walker Equations

To reduce the variance associated with the modified Yule-Walker equation estimator described above, it is possible to utilise more of the available equations. Since for an ARMA(p, q) process

$$R_{xx}[k] = \sum_{i=1}^{p} a[i] R_{xx}[k-i] \qquad k \geq q+1 \tag{18}$$

the choice of only using p equations in (9) is arbitrary. The motivation for extending this number of equations is that the autocorrelation function contains information beyond

the p lags used in equation (9) [4]. Assuming that the highest lag of the autocorrelation function which may be estimated is M, consider the following set of equations,

$$\begin{bmatrix} R_{xx}[q+1] \\ R_{xx}[q+2] \\ \vdots \\ R_{xx}[M] \end{bmatrix} = - \begin{bmatrix} R_{xx}[q] & R_{xx}[q-1] & \dots & R_{xx}[q-p+1] \\ R_{xx}[q+1] & R_{xx}[q] & \dots & R_{xx}[q-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}[M-1] & R_{xx}[M-2] & \dots & R_{xx}[M-p] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix}, \quad (19)$$

or in matrix form

$$\mathbf{r} = -\mathbf{R}\mathbf{a}. \quad (20)$$

With $\mathbf{R}$ of dimension $(M-q) \times p$, and assuming $M - q > p$, there will be more equations than unknowns. Also, since the actual autocorrelation is replaced by an estimate, allowance must be made for estimation error. If equation (20) is re-written as

$$\hat{\mathbf{r}} = -\hat{\mathbf{R}}\mathbf{a} + \mathbf{e}, \quad (21)$$

then a least-squares solution for the AR parameters [13, 18] is given by,

$$\hat{\mathbf{a}} = -\left(\hat{\mathbf{R}}^T\hat{\mathbf{R}}\right)^{-1}\hat{\mathbf{R}}^T\hat{\mathbf{r}}. \quad (22)$$

The estimator is asymptotically unbiased and has variance which usually decreases as $M - q$ (the number of equations) increases, provided $N \gg M - q$.

To determine the MA parameters and white noise variance estimate, Durbin's method may be used (as described above), again after filtering with the estimate of $A(z)$.

4.2 Mayne Firoozan Method

The Mayne-Firoozan method of ARMA parameter estimation [17] fits a high order AR model to the data, then filters the data with the inverse of this model as a means of estimating the white noise process. This filtered data are then used in conjunction with the original data to determine an ARMA input/output model via a least squares estimation.

The process consists of 5 steps,

1. Determine $\hat{\mathcal{A}}(z)$, the value of $\mathcal{A}(z)$ which minimises

$$\sum_{k=1}^{N} [\mathcal{A}(z)y[k]]^2 = \sum_{k=1}^{N} [y[k] + \alpha_1 y[k-1] + \ldots + \alpha_M y[k-M]]^2, \quad (23)$$

   where $M \gg p$.

2. Determine the residual sequence $e[k] : k = 1, 2, \ldots, N$ using

$$e[k] = \hat{\mathcal{A}}(z)y[k]. \quad (24)$$

3. Determine $\hat{A}^1(z)$ and $\hat{B}^1(z)$, the values of $A(z)$ and $B(z)$ which minimise

$$\sum_{k=1}^{N}[A(z)y[k]-B(z)e[k]] = \sum_{k=1}^{N}[(y[k]-e[k])+(a_1,\ldots,a_p,b_1,\ldots b_p) \times(y[k-1],\ldots,y[k-p]e[k-1],\ldots,e[k-p])]^2 . \quad (25)$$

4. Determine the filtered sequences $\{\tilde{y}\}$ and $\{\tilde{e}\}$ using

$$\begin{aligned}\hat{B}(z)\tilde{y}[k] &= y[k],\\ \hat{B}(z)\tilde{e}[k] &= e[k].\end{aligned} \quad (26)$$

5. Determine $\hat{A}^2(z)$ and $\hat{B}^2(z)$, the values of $A(z)$ and $B(z)$ which minimise

$$\sum_{k=1}^{N}[A(z)\tilde{y}[k]-B(z)\tilde{e}[k]] = \sum_{k=1}^{N}[(\tilde{y}[k]-\tilde{e}[k])+(a_1,\ldots,a_p,b_1,\ldots b_p) \times(\tilde{y}[k-1],\ldots,\tilde{y}[k-p]\tilde{e}[k-1],\ldots,\tilde{e}[k-p])]^2 . \quad (27)$$

Steps 1, 3 and 5 are accomplished using a linear least squares estimate. Steps 4 and 5 are added to improve the consistency of the parameter estimates.

4.3 Stochastic Recursive Least Squares

The recursive least squares algorithm is used widely in adaptive control [1, 2, 12] and adaptive filtering [8], as a method of continuously updating model parameters as new data are received. In both these fields, however, access is available to the system (or filter) inputs and outputs. As noted previously, in vibration analysis there is generally no input signal available, thus the least squares algorithm requires a slight modification [12] to produce a stochastic recursive least squares algorithm. The algorithm is summarised below, and a detailed derivation is presented in Appendix A.

The parameterised model is written in the form

$$y[n]=\varphi_1[n]\theta_1+\varphi_2[n]\theta_2+\ldots+\varphi_M[n]\theta_M, \quad (28)$$

where the model parameters are,

$$\theta=[a_1a_2\ldots a_pb_1b_2\ldots b_p], \quad (29)$$

ie., $M=2p$. The system outputs, $y[n]$, and estimated inputs, $v[n]$ are augmented to form the array,

$$\varphi^T[n]=[y[n-1]y[n-2]\ldots y[n-p]v[n-1]v[n-2]\ldots v[n-p]]. \quad (30)$$

The array of outputs used in the above model up to time n is then

$$Y[n]=\begin{bmatrix} y[1]\\ y[2]\\ \vdots\\ y[n]\end{bmatrix}, \quad (31)$$

and the model error is defined as,

$$\begin{aligned} e[n] &= y[n] - \hat{y}[n] \\ &= y[n] - \varphi^T[n]\hat{\theta}, \end{aligned} \tag{32}$$

where $\hat{y}[n]$ us the estimate of $y[n]$ based on the estimated parameters $\hat{\theta}$. The array of errors may be defined as

$$E[n] = \begin{bmatrix} e[1] \\ e[2] \\ \vdots \\ e[n] \end{bmatrix} = Y[n] - \phi[n]\hat{\theta}, \tag{33}$$

where,

$$\phi[n] = \begin{bmatrix} \varphi^T[1] \\ \varphi^T[2] \\ \vdots\varphi^T[n] \end{bmatrix} \text{ with } \varphi[n] = \begin{bmatrix} \varphi_1[n] \\ \varphi_2[n] \\ \vdots \\ \varphi_p[n] \end{bmatrix}. \tag{34}$$

Or in terms of the outputs and estimated inputs,

$$\phi[n] = \begin{bmatrix} -y[0] & -y[-1] & \dots & -y[1-p] & v[0] & v[-1] & \dots & v[1-p] \\ -y[1] & -y[0] & \dots & -y[2-p] & v[1] & v[0] & \dots & v[2-p] \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -y[n-1] & -y[n-2] & \dots & -y[n-p] & v[n-1] & v[n-2] & \dots & v[n-p] \end{bmatrix}. \tag{35}$$

For simplicity in the algorithm, define the matrix

$$P[n] = \left[\phi^T[n]\ \phi[n]\right]^{-1}. \tag{36}$$

The least squares recursive algorithm for estimating the parameter vector θ is

1. Determine $K[n]$ (refer to Appendix A for details),

$$K[n] = P[n-1]\varphi[n]\left[I + \varphi^T[n]P[n-1]\varphi[n]\right]^{-1}, \tag{37}$$

where I is the identity matrix. For a single input single output system, this reduces to

$$K[n] = \frac{P[n-1]\varphi[n]}{1 + \varphi^T[n]P[n-1]\varphi[n]}. \tag{38}$$

The matrix $K[n]$ is often referred to as the Kalman gain matrix [8].

2. Update the estimated parameter vector $\hat{\theta}$

$$\hat{\theta}[n] = \hat{\theta}[n-1] + K[n]\left[y[n] - \varphi^T[n]\hat{\theta}[n-1]\right]. \tag{39}$$

3. Update the matrix $P[n]$

$$P[n] = [I - K[n]\varphi[n]]\, P[n-1]. \tag{40}$$

The procedure is initialised by setting

$$P[0] = \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \alpha \end{bmatrix} = \alpha I, \tag{41}$$

for some large α.

The above algorithm assumes knowledge of the input sequence, which is unavailable for a stochastic model. However, since the prediction error for a stochastic model is analogous to the Gaussian distributed input sequence, it may be used as an estimate of the input sequence,

$$v[n] = e[n] = y[n] - \varphi^T[n]\hat{\theta}[n-1]. \tag{42}$$

To initialise, the error sequence is assumed zero for $n < 0$ ie., $v[n] = \mathbf{0}$. Additionally, the initial parameter vector is also set to zero, $\hat{\theta}[0] = \mathbf{0}$.

5 Fault Detection Condition Indices

In order to determine the presence or absence of a fault, some signal feature or condition index must be determined. These are generally a numerical value derived from, say, the statistics of the vibration signal. Empirical analysis and experience allow ranges of these values for which acceptable and unacceptable operation may be defined.

5.1 Current Methods

At AMRL, the use of synchronous signal averaging has been fundamental to the ability to detect faults on individual gears within a gearbox. The approach used is to average segments of the vibration signal which correspond to one shaft rotation. Such ensemble averaging allows signal frequencies which are synchronous with the shaft period for that gear to be isolated from all other signals.

The signal average contains a large amount of information which is redundant to the aim of fault detection. This is essentially contained in the gear meshing frequency and its harmonics, as well as a smaller amount at the fundamental shaft rotation frequency, and its modulation with the meshing harmonics. The latter may be due to imbalances in shaft rotation (perhaps caused by shaft misalignment). However, this once-per-rev component may be a vital factor once localised faults propagate sufficiently as shown by Forrester [6]. For example, a defect on a single gear tooth will have a prominent once-per-rev component each time it comes into mesh, provided the effect of this fault exceeds that due to shaft imbalances. Thus there is a trade-off in removing the once-per-rev component.

The information is redundant in terms of fault detection since it is relevant to the gear meshing only, and contains no information relevant to the fault. Hence it may be removed as a means of enhancing sensitivity to faults.

Removal of the tooth meshing harmonics, as well as the once-per-rev and possibly twice-per-rev modulation of these, produces a *residual signal average*. Clearly this residual signal average would be more useful for fault detection analysis, and some fault indices used at AMRL are based on this premise.

Current condition indices used at AMRL [3, 5] have been found by experience to be useful indicators of various types of gearbox faults. The indices are of two basic types: (1) general and (2) local. General indices are sensitive to general faults, such as wear on all teeth; whereas local indices are sensitive to localised perturbations, such as changes in the stiffness of a single tooth on a gear. The indices used at AMRL are summarised below.

B1 Amplitude of the shaft frequency.

A1 Amplitude of 2 $\times$ shaft frequency.

G1 RMS of the signal average.

G2 RMS of the residual signal average divided by the RMS of the signal average.

G3 The peak level of the envelope of the narrow band around a mesh harmonic divided by the RMS of the mesh harmonic.

L1 Kurtosis of the residual signal.

L2 Kurtosis of the envelope of the narrow band around a mesh harmonic.

L3 Kurtosis of the instantaneous frequency of the modulated signal.

5.2 Methods for Diagnosis

As described above the use of synchronous signal averaging and removal of gear mesh harmonics allows the removal of a large amount of redundant information (at least in terms of fault detection). The signal averaging process enables frequencies relevant to the particular gear of interest to be extracted from those due to other gears (provided, of course, other gears do not have rotation frequencies and harmonics common to that gear). Further to this, the *residual signal average* removes the gear meshing harmonics which are of little interest to fault detection.

With this in mind, and the use of ARMA modelling as a means of accounting for load variation, the residual signal obtained from the model appears intuitively to offer the most promise as a means of fault detection. The residual signal, in this case, is just the difference between the measured vibration signal, and the one-step-ahead prediction based on the estimated ARMA model,

$$\begin{aligned} \varepsilon[n] &= x[n] - \hat{x}[n] \\ &= x[n] + \sum_{k=1}^{p} \hat{a}[k]x[n-k] - \sum_{k=1}^{p} \hat{b}[k]v[n-k] \end{aligned} \tag{43}$$

It should be stressed that the residual signal here is not the same as the residual signal average[1].

Since the ARMA model attempts to model the dynamics of the process, the residual signal is ideally a Gaussian distributed random process of variance σ^2, ie., it is equivalent to the input of the ARMA model. To clarify this, the ARMA model is defined as a Gaussian process passed through a linear filter of the form $B(z)/A(z)$. Using this model to predict the output, the error must be equivalent to the noise input since this is the part of the signal which cannot be predicted.

Once a fault is introduced, the structure of the process model changes. Thus using a predictor based on a fault-free process will change the residual signal from Gaussian noise to coloured (filtered) noise since there will be un-modelled dynamics due to the introduction of the fault.

From a different perspective, but conceptually the same, the residual signal is akin to the process of synchronous signal averaging and removal of the gear mesh harmonics as described above. The most notable difference is that the latter uses ensemble averaging whereas the former does not. The residual signal consists of the original signal minus the identified process dynamics. Hence removal of the redundant information is similarly achieved.

The residual signal, then, provides the most obvious means of detecting the presence of a fault. Statistical properties and the power spectral density offer the best means of extracting information, from this signal, which may be used to detect faults.

Coherent signal averaging offers another advantage hitherto unmentioned. That is it provides a convenient description of the signal from a particular gear. In a crude sense, ensemble averaging compresses the information recorded over many minutes into a short data record, which describes the rotation of a single gear, and this is primarily what we are interested in. Application of synchronous signal averaging to the residual signal from the ARMA model will then enable extraction of information pertinent to the rotation of a single gear, particularly if localised faults are present, since (ideally) all other information has been removed. In this case, current fault indices based on the residual signal average may be use on the signal average derived from the ARMA model residual signal.

The use of ARMA modelling, then, enables load variation to be accounted for, and the residual of the model provides the means of detecting faults.

6 Simulations

In this section a simple ARMA time-series is generated to simulate a linear process. To simulate varying load conditions, the poles and zeros of the model will be altered in steps. To simulate the action of a fault, a second ARMA time-series is generated, and

[1]The residual signal average is the ensemble average of the vibration signal with the meshing harmonics removed. The residual signal in this case is the output of a prediction error filter, and no ensemble averaging is applied.

added to the first for a short period of time. Since the dynamics of the fault are also likely to change with varying load, the poles and zeros of the fault signal are also varied with load.

The following simulations are intended only as an illustration of how load variation may be accounted for, and how fault detection is possible. Further work using data recorded from the AMRL gear test rig will appear in a subsequent report.

6.1 General Time Series Simulation

To generate ARMA time-series examples with a known transfer function, it is necessary to separate the AR and MA sections of the model. Consider the ARMA time-series model of equation (5)

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + \sum_{k=1}^{q} b[k]v[n-k], \tag{44}$$

where $\{v[k] : k = 1, 2, \ldots, p\}$ is a Gaussian distributed white noise process. Since the AR section requires knowledge of $x[n]$ for $n < 0$ (and this is clearly not available) it is necessary to determine appropriate initial conditions. This is achieved by setting [13]

$$y[n] = \begin{cases} \sqrt{\rho_0}v[0] & \text{for } n = 0 \\ -\sum_{l=1}^{k} a_n[l]y[n-l] + \sqrt{\rho_n}v[n] & \text{for } n = 1, 2, \ldots, p-1 \end{cases}, \tag{45}$$

where $a_i[j]$ is the ith prediction coefficient for the jth order model, and ρ_j is the prediction error power for the jth order model. This requires an additional $p-1$ sets of prediction coefficients (for the predictors of order $1, 2, \ldots, p-1$) which may be derived from the pth order AR model by [13]

$$\begin{aligned} a_{k-1}[i] &= \frac{a_k[i] - a_k[k]a_k[k-i]}{1 - |a_k[k]|^2} \qquad \text{for } i = 1, 2, \ldots, k-1 \quad k = p-1, p-2, \ldots, 1 \\ \rho_{k-1} &= \frac{\rho_k}{1 - |a_k[k]|^2} \qquad \text{for } k = p-1, p-2, \ldots, 1. \end{aligned} \tag{46}$$

With the above initial conditions set, the remaining data for the AR model may be generated using

$$y[n] = -\sum_{k=1}^{p} a[n]y[n-k] + \sqrt{\rho_p}v[n] \qquad n = p, p+1, \ldots, N+q, \tag{47}$$

where N is the total number of time-series samples required.

Filtering the AR time-series $\{y[n] : n = 1, 2, \ldots, N\}$ with a finite impulse response (FIR) filter consisting of coefficients equal to the MA coefficients will then produce the required ARMA time-series,

$$x[n] = \sum_{k=1}^{q} b[k]y[n-k] \qquad n = q, q+1, \ldots, N+q. \tag{48}$$

Note that a delay of q samples is required to ensure the filter is initialised.

This ARMA time-series is completely described by its ARMA parameters $\{a[k] : k = 1, 2, \ldots, p\}$ and $\{b[k] : k = 1, 2, \ldots, q\}$ as well as the prediction error power ρ_p.

6.2 System Model and Fault Simulation

Since the fundamental frequency components of the vibration signal due to gear meshing are directly related to the shaft rotation frequency, then load variation will not change these. Hence, the argument (angle) of the model poles and zeros in the $\mathcal{Z}$-domain will remain the same. It is, however, possible that the damping of the poles and zeros (distance from the origin) may change with load.

The simple system model, for illustrating how load variation may be accounted for, consists of two pairs of complex conjugate poles, and a double real zero, as shown in figures 4 and 5. Note that the two figures show the initial and final load positions. Also, the double zero appears only once on the real axis. To simulate load variation the poles and zeros are moved from the initial position (figure 4) to the final position (figure 5) in five steps.

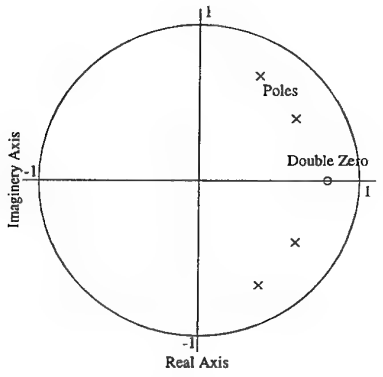

Figure 4: Initial system poles and zeros.

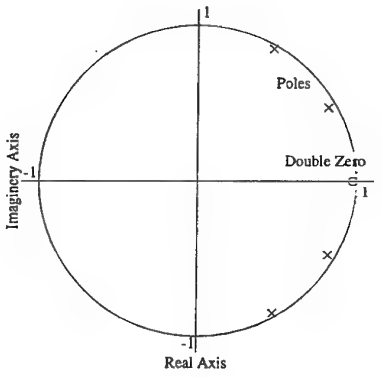

Figure 5: Final system poles and zeros.

Previous analysis of vibration signal from helicopter transmission systems with damaged teeth [5] has shown that the fault appears, predominantly, at the second and higher order harmonics of the tooth-meshing frequency. To exaggerate this, for illustration purposes, the fault signal is simulated as a pair of complex conjugate poles, and a single zero in the left hand plane of the $\mathcal{Z}$-domain. These are shown in figures 6 and 7. Again load variation is simulated by moving the poles and zero away from the origin towards the unit circle in five steps. In addition, a low signal-to-noise ratio is achieved by setting the white noise variance used to generate the fault time-series to one tenth that of the system signal.

The spectra of the system and fault processes are shown in figures 8–11. Note that the magnitude of the system spectra are significantly greater than that of the fault spectra.

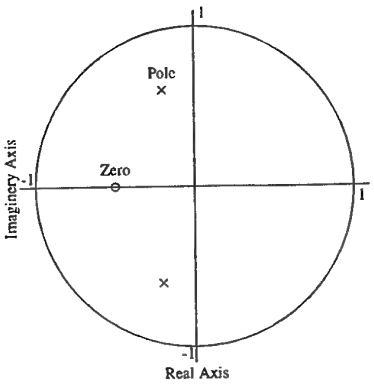

Figure 6: Initial fault poles and zero.

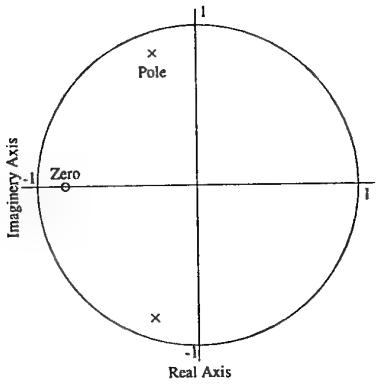

Figure 7: Final fault poles and zero.

These spectra are determined from the actual ARMA parameters using,

$$|P_{ARMA}[f]| = \frac{\hat{\sigma}^2 \sum_{k=0}^{q} b[k] e^{-j2\pi fk}}{1 + \sum_{k=1}^{p} a[k] e^{-j2\pi fk}}. \tag{49}$$

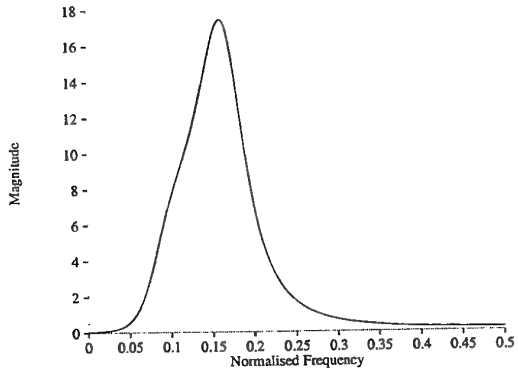

Figure 8: Initial system ARMA spectra.

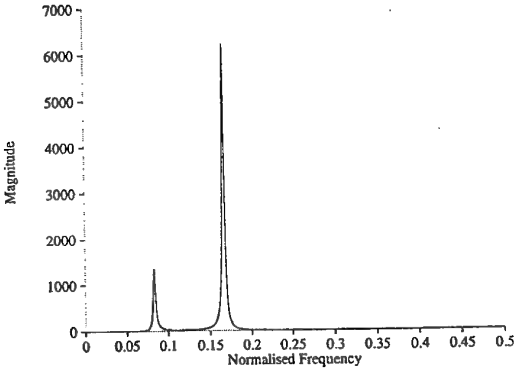

Figure 9: Final system ARMA spectra.

Figures 12 and 13 show 30,000 samples of the system ARMA simulation to which has been added 10,000 samples of the fault signal in the sample range 10,000–20,000. From these figures the location of the fault is not apparent in either of the simulations.

Since the kurtosis provides a measure of localised faults, a series of impulses with a magnitude of 10.0 is added to the fault signal. The impulses are spaced 100 samples apart. Time-series simulations for the initial and final load conditions are shown in figures 14 and 15. Again the location of the fault is not distinguishable, despite the additional impulses with a magnitude of 10.0.

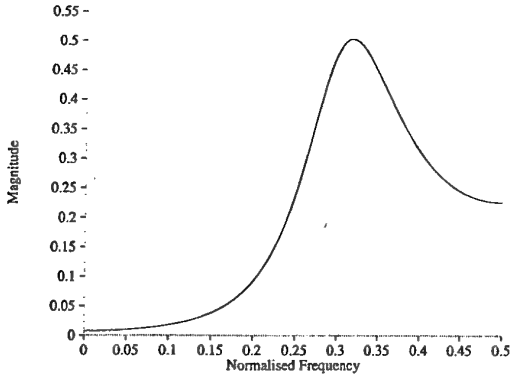

Figure 10: Initial fault ARMA spectra.

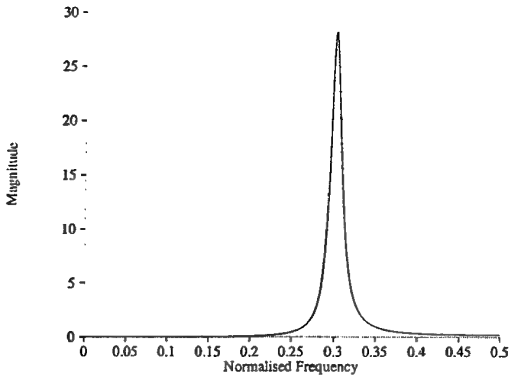

Figure 11: Final fault ARMA spectra.

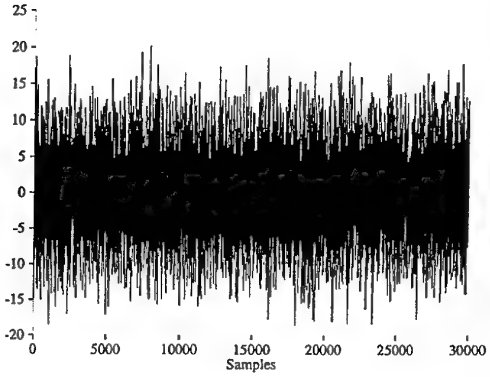

Figure 12: Initial system and fault signal time-series.

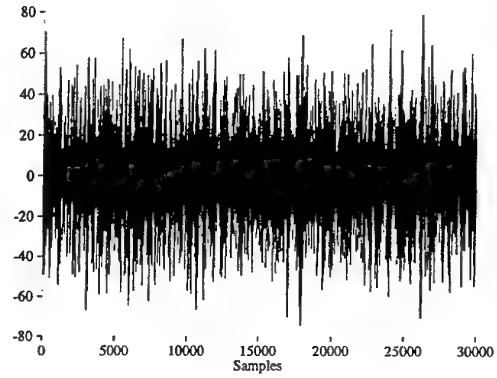

Figure 13: Final system and fault signal time-series.

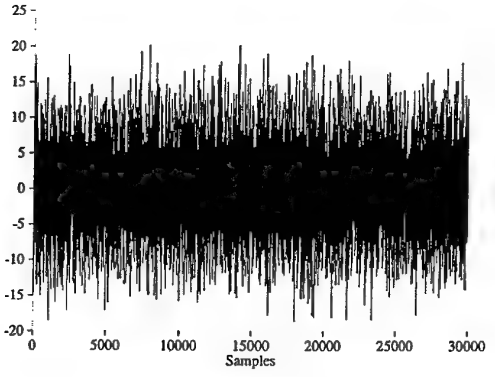

Figure 14: Initial system and fault signal time-series with impulses.

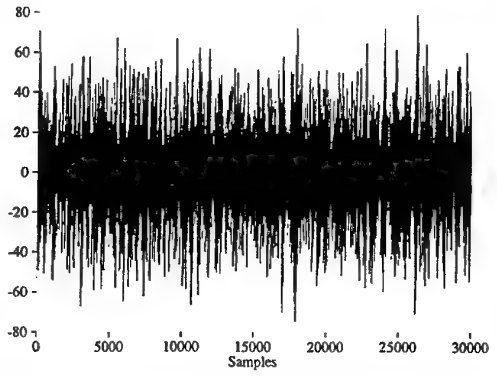

Figure 15: Final system and fault signal time-series with impulses.

6.3 Fault Detection

To test both general and local fault detection indices, the variance (equivalent to the RMS) and kurtosis of the residual signal are to be determined. To generate the residual signal, a prediction error filter is formed using the ARMA model. A diagram is shown in figure 16.

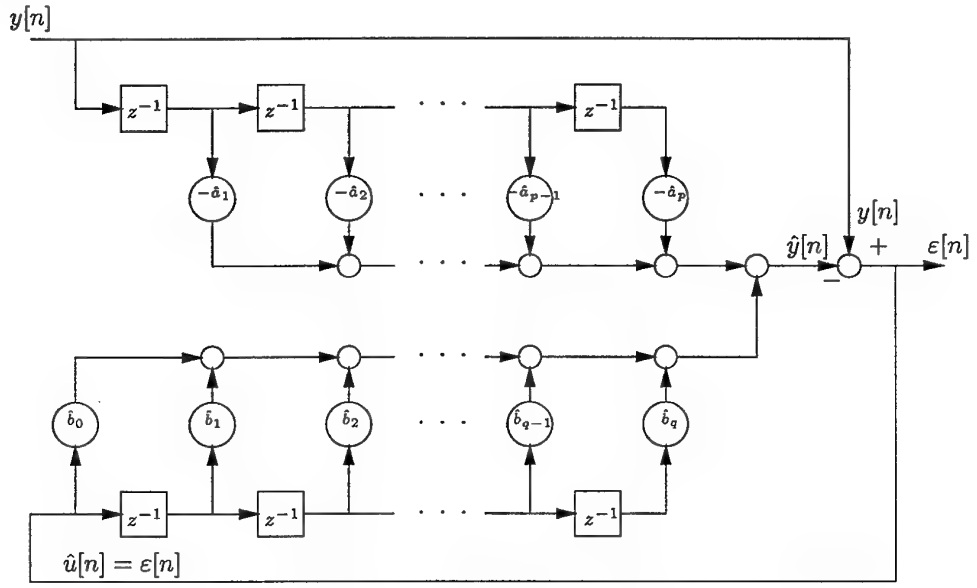

Figure 16: Prediction error filter block diagram.

For the data generated above in Section 6.2 an ARMA(2,2) model is estimated, and a prediction error filter, as shown in figure 16, is formed using the parameter estimates. Figures 17 and 18 show the residual signals determined from the initial and final fault signal time-series shown in figures 12 and 13. In figure 17 a small increase in signal magnitude can be seen from 10,000–20,000 samples (the location of the fault), however, the increase is far more evident in figure 18. This suggests that some measure of the signal magnitude will be suitable for detecting a fault. The reason that the increase is greater for the final fault signal is that the poles and zeros of the fault signal model are closer to the unit circle. This has the effect of creating a sharper notch in the power spectral density, and increasing the magnitude of the fault signal (compare figures 10 and 11, and note the differences in the magnitude scales).

Applying the same prediction error filter to the time-series with impulses added to the fault signals give the results shown in figures 19 and 20. In this case both the initial and final fault signal residual signals show a significant increase in magnitude at the location of the fault. The greater discrimination here is essentially due to the impulses, which are not accounted for by the ARMA model (since they are part of the fault signal).

Figures 21–25 show the variance of the residual signal for each of the fault signal time-series as described above. The variance was determined by averaging over 1,000 samples

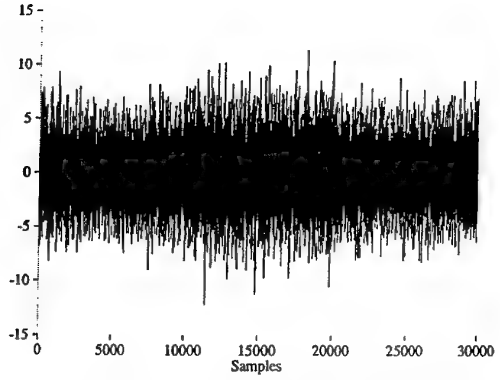

Figure 17: Residual signal from initial fault signal time-series.

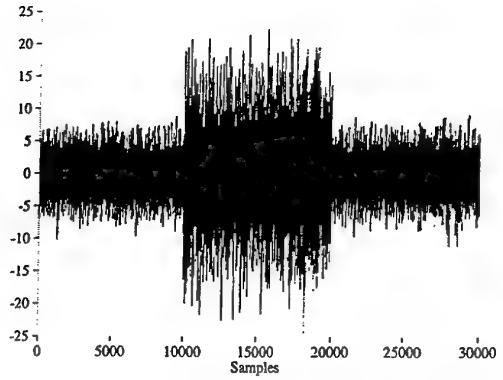

Figure 18: Residual signal from final fault signal time-series.

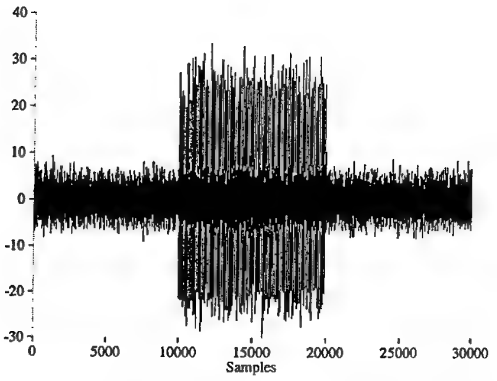

Figure 19: Residual signal from initial fault signal time-series with impulses.

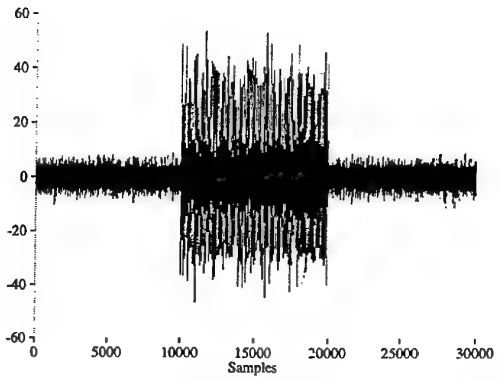

Figure 20: Residual signal from final fault signal time-series with impulses.

using a sliding block approach. Here, the variance is determined over the block of 1,000 samples. Then as each new sample is added to the start of the block, one sample is discarded from the end, and the variance is determined for the new data block. This gives the variance over the averaging time (of 1,000 samples) as a function of samples (or time).

Note that despite there being little discrimination between normal and fault conditions in the residual signal of figure 17 (for the initial fault signal), figure 21 shows a clear distinction from 10,000–20,000 samples. This is a consequence of the averaging process used in determining the variance.

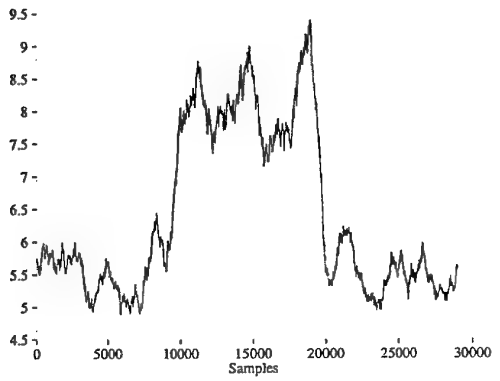

Figure 21: Variance of residual signal from initial fault signal time-series.

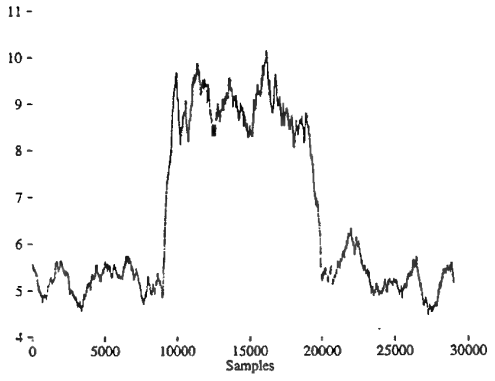

Figure 22: Variance of residual signal from second fault signal time-series.

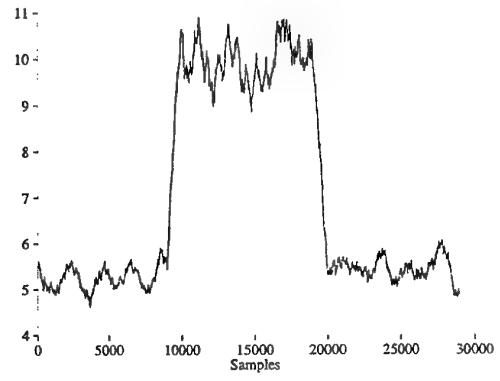

Figure 23: Variance of residual signal from third fault signal time-series.

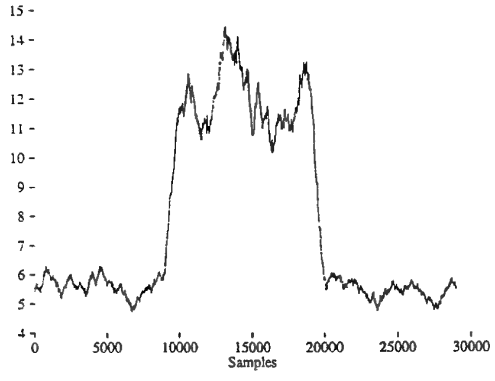

Figure 24: Variance of residual signal from fourth fault signal time-series.

Figures 26–30 show the variance determined for the residual of the fault signal time-series with impulses added. It is evident that the addition of impulses allows a greater discrimination between normal and fault conditions. This should be understood by comparing the residual signals of figures 17 and 19, which show that the impulses lead to a significant increase in the magnitude of the residual signal when the fault is present.

Figures 31–35 show the kurtosis of the residual signal for each of the fault signal time-series described previously. A similar sliding block approach, as described previously, is

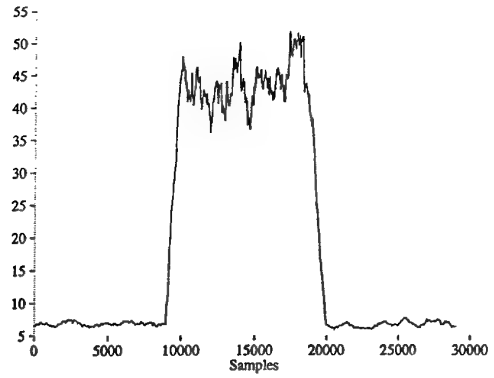

Figure 25: Variance of residual signal from final fault signal time-series.

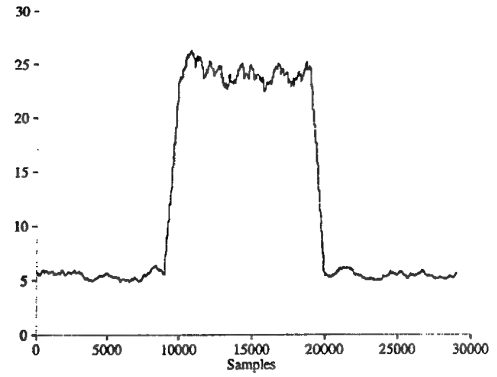

Figure 26: Variance of residual signal from initial fault signal time-series with impulses.

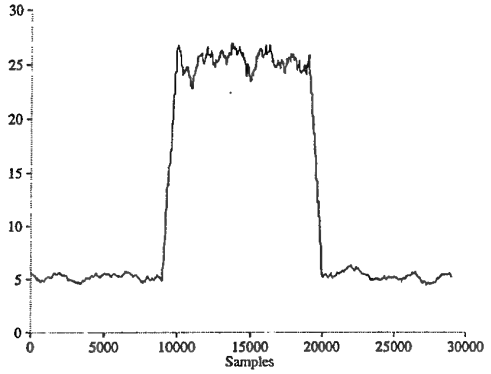

Figure 27: Variance of residual signal from second fault signal time-series with impulses.

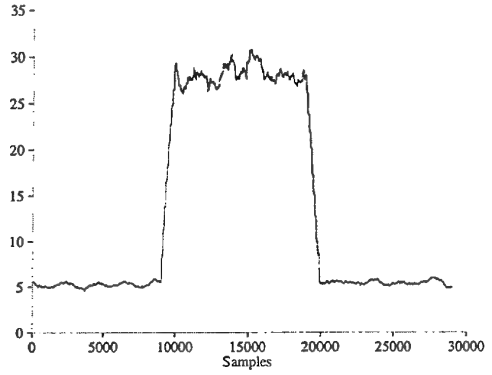

Figure 28: Variance of residual signal from third fault signal time-series with impulses.

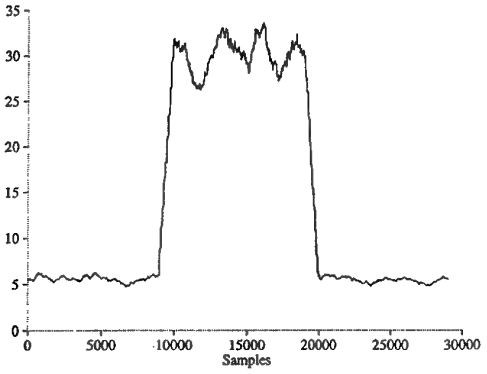

Figure 29: Variance of residual signal from fourth fault signal time-series with impulses.

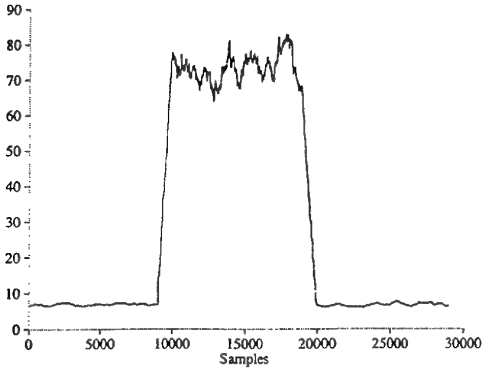

Figure 30: Variance of residual signal from final fault signal time-series with impulses.

also used with averaging over 1,000 samples. In each of these figures it is not possible to determine the location of the fault, since the kurtosis appears to be similar throughout each record. There is, however, a distinct increase in the kurtosis at the beginning and end of the fault section in figure 35. This can be attributed to the transition from normal to fault condition, and thence from fault to normal condition. This feature implies that the kurtosis is able to detect transient phenomena in the residual signal, provided it is of significant magnitude. The transition in the residual signal can be seen clearly in figure 18, where the fault signal from 10,000–20,000 samples has a much greater magnitude than the normal signal either side of this.

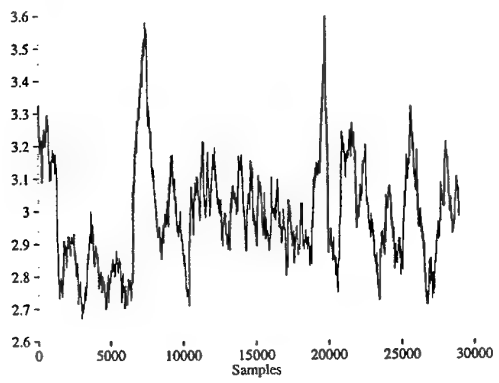

Figure 31: Kurtosis of residual signal from initial fault signal time-series.

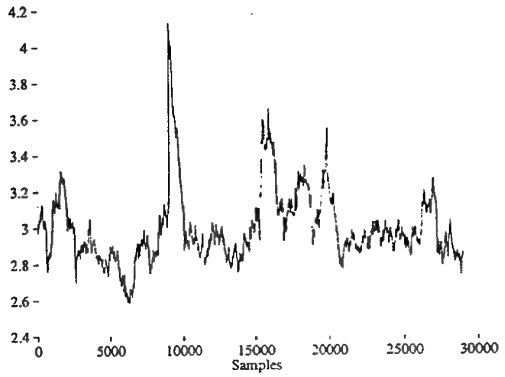

Figure 32: Kurtosis of residual signal from second fault signal time-series.

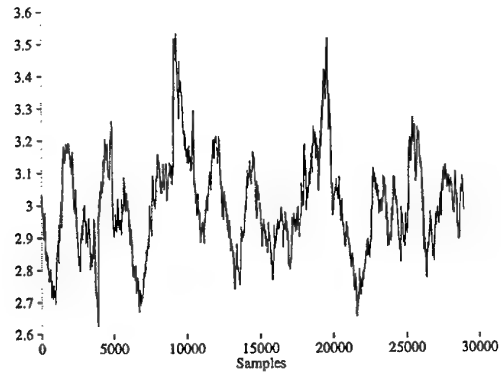

Figure 33: Kurtosis of residual signal from third fault signal time-series.

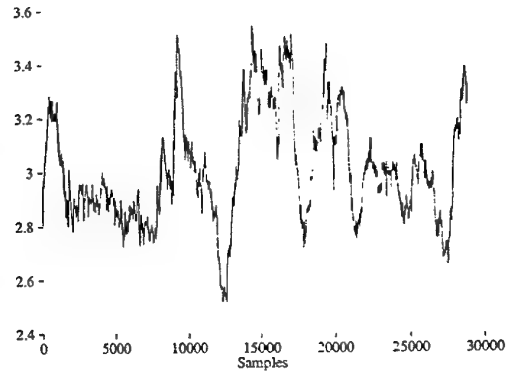

Figure 34: Kurtosis of residual signal from fourth fault signal time-series.

Figures 36–40 show the kurtosis determined for the residual of the fault signal time-series with impulses added. In contrast to figures 31–35 the location of the fault can be seen clearly in the sample range 10,000–20,000. Also noticeable are sharp peaks at the beginning and end of the fault section. This shows that the kurtosis is suitable for determining faults which are impulsive, as well as being able to detect the transition as the fault signal is added to the normal signal.

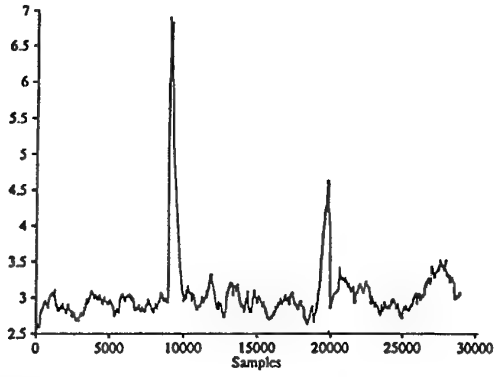

Figure 35: Kurtosis of residual signal from final fault signal time-series.

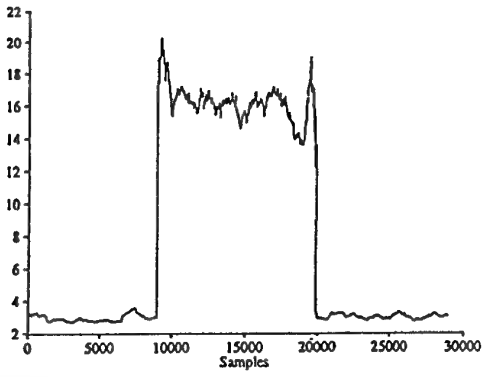

Figure 36: Kurtosis of residual signal from initial fault signal time-series with impulses.

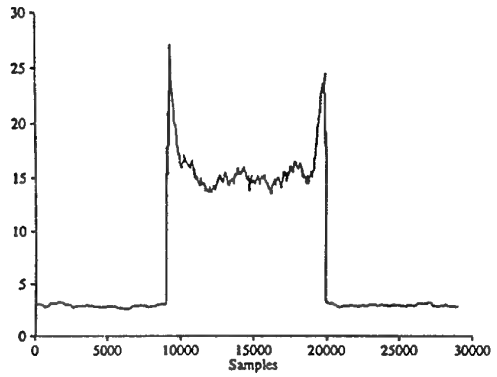

Figure 37: Kurtosis of residual signal from second fault signal time-series with impulses.

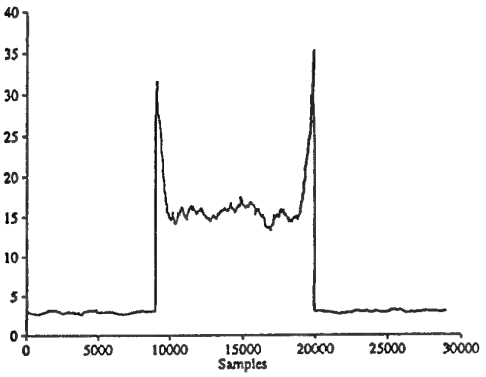

Figure 38: Kurtosis of residual signal from third fault signal time-series with impulses.

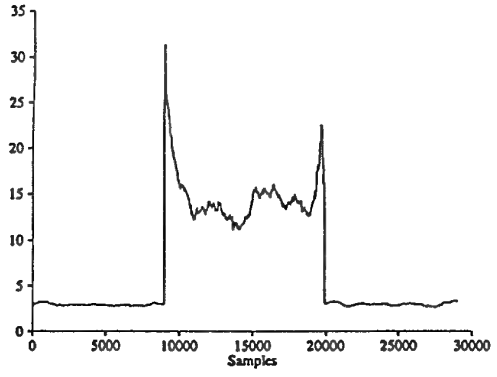

Figure 39: Kurtosis of residual signal from fourth fault signal time-series with impulses.

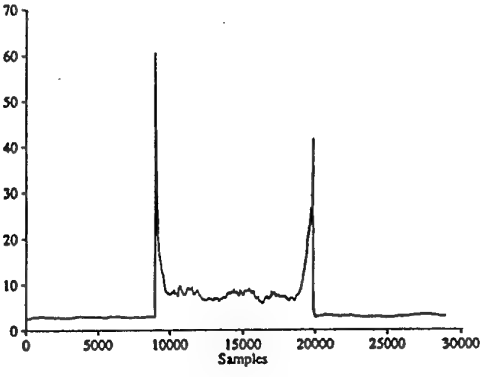

Figure 40: Kurtosis of residual signal from final fault signal time-series with impulses.

7 Further Requirements

This report is intended to review signal processing methods which may be suitable for accounting for variable load conditions when analysing gearbox vibration signals for fault detection. The application of these methods requires much further study.

Before applying these methods, however, if is necessary to make a comprehensive study of vibration signals under varying load conditions. It will be necessary to achieve this using signals which are known to be fault-free. It is important to gain as much information as possible about "normal" condition signals particularly if little, or no, information is to be assumed regarding the fault. The main reason for this is that it is not possible to characterise all types of faults, since there are perhaps many faults for which we have no knowledge of [23].

The methods described will have to be applied to actual vibration signals to determine the suitability of the types of models presented. Following this, it will be necessary to determine how effective they are in detecting faults. In particular, the time between detection and catastrophic failure will be a most important indicator here.

Previously it was suggested that model parameters will need to be determined at several different load settings. Subsequently, the analysis for fault detection will also have to be achieved at different load settings. However, depending on the variation of model parameters, it may be possible to interpolate model parameters between these fixed load points to achieve a more practical system, able to be used at any load.

An obvious extension would be to apply synchronous signal averaging to the ARMA residual signal. This has two possible benefits namely: (1) it will be possible to determine whether the residual signal contains periodic components at the meshing frequencies – which ideally should have been removed, and (2) as a means of data reduction. This latter point will be of benefit if on-board hardware is to be used, in which case data storage will be at a premium.

8 Conclusions

This report introduced linear modelling techniques as a means of accounting for varying load conditions experienced in helicopter transmission systems. Several methods of ARMA parameter estimation were outlined which enable modelling of the vibration signals recorded from transmission systems. With the ultimate aim being the detection of faults under varying load conditions, it was suggested that a prediction error filter be used to generate a residual signal. Subsequent analysis of this residual signal will then enable detection of faults.

A comparison was made between the residual signal derived from the prediction error filter, and the residual signal average in common use by AMRL. It was proposed that the former is able to remove frequency components which are inherent in the vibration signal (including those synchronous with the shaft rotation of the gear of concern, and other gears in the gearbox), whereas, synchronous signal averaging is incapable of removing any signal

component which is synchronous with the shaft rotation. The removal of the meshing harmonics and once-per-rev modulations is a means of removing information redundant to fault detection, from the signal average. However, as noted previously, this may also remove information specific to the fault. For example, the once-per-rev component will be evident if a localised gear fault (such as a tooth crack) is present, in addition to a similar component due to shaft imbalances. The residual signal derived from the prediction error filter does not suffer this disadvantage since the model is estimated from a vibration signal which is known to be fault free. Thus the residual signal should only contain information which is relevant to the fault.

Simple simulations were used to illustrate the ability of the technique to detect a fault under varying load conditions. A simulated fault signal was added to a simulated vibration signal for a short period. It was found that the residual signal derived from the prediction error filter had an increase in power during the time the fault signal was applied. Two indices were determined from the residual signal; the (1) variance and (2) kurtosis. It was found that the variance was sensitive to changes in signal magnitude, and the kurtosis to transients in the residual signal.

It was suggested that synchronous signal averaging of the residual signal from the prediction error filter would be a useful means of data reduction, as well as a check on the effectiveness of the methods in removing information redundant to fault detection (such as the gear meshing harmonics).

It was noted that further work is required in analysing gear rig vibration signals to determine the applicability of the modelling techniques.

Appendix A
Stochastic Recursive Least Squares Derivation

This appendix gives a derivation of the recursive least squares algorithm adapted for stochastic signals. It is drawn from several sources [1, 2, 12, 8] in an attempt to provide a more concise and simplified derivation.

The parameterised model is written in the form

$$y[n] = \varphi_1[n]\theta_1 + \varphi_2[n]\theta_2 + \ldots + \varphi_M[n]\theta_M, \tag{A1}$$

where the model parameters are,

$$\theta = [a_1 a_2 \ldots a_p b_1 b_2 \ldots b_p], \tag{A2}$$

ie., $M = 2p$. The system outputs, $y[n]$, and estimated inputs, $v[n]$ are augmented to form the array,

$$\varphi^T[n] = [y[n-1]y[n-2]\ldots y[n-p]v[n-1]v[n-2]\ldots v[n-p]]. \tag{A3}$$

The array of outputs used in the above model up to time n is then

$$Y[n] = \begin{bmatrix} y[1] \\ y[2] \\ \vdots \\ y[n] \end{bmatrix}, \tag{A4}$$

and the model error is defined as,

$$\begin{aligned} e[n] &= y[n] - \hat{y}[n] \\ &= y[n] - \varphi^T[n]\hat{\theta}, \end{aligned} \tag{A5}$$

where $\hat{y}[n]$ us the estimate of $y[n]$ based on the estimated parameters $\hat{\theta}$. The array of errors may be defined as

$$E[n] = \begin{bmatrix} e[1] \\ e[2] \\ \vdots \\ e[n] \end{bmatrix} = Y[n] - \phi[n]\hat{\theta}, \tag{A6}$$

where,

$$\phi[n] = \begin{bmatrix} \varphi^T[1] \\ \varphi^T[2] \\ \vdots\varphi^T[n] \end{bmatrix} \text{ with } \varphi[n] = \begin{bmatrix} \varphi_1[n] \\ \varphi_2[n] \\ \vdots \\ \varphi_p[n] \end{bmatrix}. \tag{A7}$$

Or in terms of the outputs and estimated inputs,

$$\phi[n] = \begin{bmatrix} -y[0] & -y[-1] & \dots & -y[1-p] & v[0] & v[-1] & \dots & v[1-p] \\ -y[1] & -y[0] & \dots & -y[2-p] & v[1] & v[0] & \dots & v[2-p] \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -y[n-1] & -y[n-2] & \dots & -y[n-p] & v[n-1] & v[n-2] & \dots & v[n-p] \end{bmatrix}. \quad \text{(A8)}$$

The least squares solution requires the sum of the errors squared to be minimised. The sum of the errors squared is given by (note that the time index is dropped for convenience)

$$\begin{aligned} V &= \sum_{i=1}^{n} e^2[i] = E^T E \\ &= \left[Y^T - \hat{\theta}^T \phi\right] \cdot \left[Y - \phi\hat{\theta}\right] \\ &= Y^T Y - Y^T \phi\hat{\theta} - \hat{\theta}^T \phi^T Y + \hat{\theta}^T \phi^T \phi\hat{\theta}. \end{aligned} \quad \text{(A9)}$$

To minimise V with respect to $\hat{\theta}$, differentiate w.r.t. $\hat{\theta}$ and set to zero,

$$\begin{aligned} \frac{\partial}{\partial\hat{\theta}} & \left[Y^T Y - Y^T \phi\hat{\theta} - \hat{\theta}^T \phi^T Y + \hat{\theta}^T \phi^T \phi\hat{\theta}\right] \\ & = 0 - \phi^T Y - \phi^T Y + 2\phi^T \phi\hat{\theta} \\ & = -2\phi^T Y + 2\phi^T \phi\hat{\theta} = 0. \end{aligned} \quad \text{(A10)}$$

Now the second derivative $(\partial/\partial\hat{\theta}^T \; \partial/\partial\hat{\theta})V = 2\phi^T\phi \geq 0$, thus V is minimised. The condition for a least squares estimate of θ is then

$$\begin{aligned} 2\phi^T\hat{\theta} &= 2\phi^T Y \\ \hat{\theta} &= \left[\phi^T\phi\right]^{-1} \phi^T Y. \end{aligned} \quad \text{(A11)}$$

A recursive algorithm for estimating θ will allow continuous updating of the parameters. To achieve this, it is necessary to relate the parameters estimated at time n to those already estimated at time $n-1$. First, define the matrix

$$P[n] = \left[\phi^T[n]\phi[n]\right]^{-1}, \quad \text{(A12)}$$

and note from equation (A7) that

$$\phi[n-1] = \begin{bmatrix} \varphi^T[1] \\ \varphi^T[2] \\ \vdots \\ \varphi^T[n-1] \end{bmatrix}, \quad \text{(A13)}$$

so that

$$P[n-1] = \left[\varphi^T(1)\varphi^T[2]\dots\varphi^T[n-1]\right] \begin{bmatrix} \varphi^T[1] \\ \varphi^T[2] \\ \dots \\ \varphi^T[n-1] \end{bmatrix}, \quad \text{(A14)}$$

then

$$P^{-1}[n] = P^{-1}[n-1] + \varphi[n]\varphi^T[n]. \tag{A15}$$

Recall from equation (A11) that

$$\begin{aligned} \hat{\theta}[n] &= \left[\phi^T\phi\right]^{-1}\phi^T Y \\ &= \left[\sum_{i=1}^{n}\varphi[i]\varphi^T[i]\right]^{-1}\left[\sum_{i=1}^{n}\varphi[i]y[i]\right] \\ &= P[n]\left[\sum_{i=1}^{n}\varphi[i]y[i]\right] \\ &= P[n]\left[\sum_{i=1}^{n-1}\varphi[i]y[i]\right] + \varphi[n]y[n], \end{aligned} \tag{A16}$$

the third line of the above equation implies

$$\hat{\theta}[n-1] = P[n-1]\left[\sum_{i=1}^{n-1}\varphi[i]y[i]\right], \tag{A17}$$

or

$$\begin{aligned} \sum_{i=1}^{n-1}\varphi[i]y[i] &= P^{-1}[n-1]\hat{\theta}[n-1] \\ &= \left[P^{-1}[n] - \varphi[n]\varphi^T[n]\right]\hat{\theta}[n-1]. \end{aligned} \tag{A18}$$

Where the last line is as a result of using $P^{-1}[n-1] = P^{-1}[n] - \varphi[n]\varphi^T[n]$ from equation (A15). Using this in the last line of (A16)

$$\begin{aligned} \hat{\theta}[n] &= P[n]\left[P^{-1}[n]\hat{\theta}[n-1] - \varphi[n]\varphi^T[n]\hat{\theta}[n-1] + \varphi[n]y[n]\right] \\ &= \hat{\theta}[n-1] + P[n]\varphi[n]\left[y[n] - \varphi^T[n]\hat{\theta}[n-1]\right]. \end{aligned} \tag{A19}$$

Note that in the last line $\varphi[n]\hat{\theta}[n-1]$ is just a one-step-ahead prediction of $y[n]$ based on the parameters estimated at time $n-1$. The expression in the brackets is then the prediction error $e[n] = y[n] - \hat{y}[n]$. The above equation can then be re-written as

$$\hat{\theta}[n] = \hat{\theta}[n-1] + K[n]e[n], \tag{A20}$$

which is appropriately recursive. It remains to determine a recursion equation for $P[n]$. Recall from equation (A15) that

$$P[n] = \left[P[n-1] + \varphi[n]\varphi^T[n]\right]^{-1}. \tag{A21}$$

Using the matrix inversion lemma [2]

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B\left[C^{-1} + DA^{-1}B\right]^{-1}DA^{-1}, \tag{A22}$$

where A, C and $[C^{-1} + DA^{-1}B]$ are required to be non-singular. Let

$$\begin{aligned} A &= P^{-1}[n-1] \\ B &= \varphi[n] \\ C &= I \qquad \text{the identity matrix} \\ D &= \varphi^T[n], \end{aligned} \tag{A23}$$

then

$$\begin{aligned} P[n] &= \left[P^{-1}[n] + \varphi[n]\varphi^T[n]\right]^{-1} \\ &= P[n-1] - P[n-1]\varphi[n]\left[I + \varphi^T[n]P[n-1]\varphi[n]\right]^{-1}\varphi^T[n]P[n-1]. \end{aligned} \tag{A24}$$

Since $K[n] = P[n]\varphi[n]$, we have

$$\begin{aligned} K[n] &= P[n-1]\varphi[n] - P[n-1]\varphi[n]\left[I + \varphi^T[n]P[n-1]\varphi[n]\right]^{-1}\varphi^T[n]P[n-1]\varphi[n] \\ &= P[n-1]\varphi[n]\left[I - \left[I + \varphi^T[n]P[n-1]\varphi[n]\right]\varphi^T[n]P[n-1]\varphi[n]\right] \\ &= P[n-1]\varphi[n]\left[I + \varphi^T[n]P[n-1]\varphi[n]\right]^{-1}, \end{aligned} \tag{A25}$$

where the last line is as a result of applying the matrix inversion lemma to the contents of the outer brackets of the line above. From equations (A24) and (A25)

$$\begin{aligned} P[n] &= P[n-1] - P[n-1]\varphi[n]\left[I + \varphi^T[n]P[n-1]\varphi[n]\right]^{-1}\varphi^T[n]P[n-1] \\ K[n] &= P[n-1]\varphi[n]\left[I + \varphi^T[n]P[n-1]\varphi[n]\right], \end{aligned} \tag{A26}$$

it follows that

$$P[n] = [I - K[n]\varphi[n]]\,P[n-1]. \tag{A27}$$

The full recursive procedure for estimating the parameter vector $\hat{\theta}$ is then,

1. Update $K[n]$

$$K[n] = P[n-1]\varphi[n]\left[I + \varphi^T[n]P[n-1]\varphi[n]\right]^{-1} \tag{A28}$$

for a single-input single-output system, this reduces to

$$K[n] = \frac{P[n-1]\varphi[n]}{1 + \varphi^T[n]P[n-1]\varphi[n]} \tag{A29}$$

2. Update $\hat{\theta}[n]$

$$\hat{\theta}[n] = \hat{\theta}[n-1] + K[n]\left[y[n] - \varphi^T[n]\hat{\theta}[n-1]\right] \tag{A30}$$

3. Update $P[n]$

$$P[n] = [I - K[n]\varphi[n]]\,P[n-1] \tag{A31}$$

The procedure is initialised be setting

$$P[0] = \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha \end{bmatrix} = \alpha I \tag{A32}$$

The above algorithm assumes knowledge of the input sequence, which is unavailable for a stochastic model. However, since the prediction error for a stochastic model is analogous to the Gaussian distributed input sequence, it may be used as an estimate of the input sequence,

$$v[n] = e[n] = y[n] - \varphi^T[n]\hat{\theta}[n-1]. \tag{A33}$$

To initialise, the error sequence is assumed zero for $n < 0$ ie., $v[n] = \mathbf{0}$. Additionally, the initial parameter vector is also set to zero, $\hat{\theta}[0] = \mathbf{0}$.

References

1. K J Astrom and Wittenmark. *Computer Controlled Systems : Theory and Design.* Prentice Hall, 1984.

2. K J Astrom and Wittenmark. *Adaptive Control.* Addison Wesley, 1989.

3. D M Blunt and B D Forrester. Health monitoring of blackhawk and seahawk main transmission using vibration analysis. In *Proceedings of the Second International Conference on Aerospace Science and Technology/Sixth Australian Aeronautical Conference*, pages 33–39, Melbourne, Australia, March 1995.

4. J A Cadzow. Spectral estimation: An overdetermined rational model equation approach. *Proceedings of the IEEE*, 70(9), 1982.

5. B D Forrester. Time-frequency analysis of helicopter transmission vibration. ARL Propulsion Report 180, 1991.

6. B D Forrester. *Advanced Vibration Analysis Techniques for Fault Detection and Diagnosis in Geared Transmission Systems.* PhD thesis, Swinburne University of Technology, 1996. submitted February.

7. Jerome H Friedman. Multivariate adaptive regression splines. *The Annuls of Statistics*, 19(1):1–141, January 1991.

8. Simon Haykin. *Adaptive Filter Theory.* Prentice Hall, 1996.

9. Robert Hecht-Nielsen. *Neurocomputing.* Addison Wesley, 1990.

10. John Hertz, Anders Krough, and Richard G. Palmer. *Introduction to the Theory of Neural Computation.* Addison Wesley, 1991.

11. Don R Hush and Bill G Horne. Progress in supervised neural networks. *IEEE Signal Processing Magazine*, pages 8–39, January 1993.

12. R Iserman. Fault diagnosis of machines via parameter estimation and knowledge processing - tutorial paper. *Automatica*, 29(4):815–835, 1993.

13. S M Kay. *Modern Spectral Estimation : Theory and Application.* Prentice Hall, 1988.

14. S M Kay and S L Marple. Spectrum analysis – a modern perspective. *Proceedings of the IEEE*, 69(11):1380–1419, 1981.

15. Richard P Lippman. An introduction to computing with neural nets. *IEEE Signal Processing Magazine*, pages 4–22, April 1987.

16. S L Marple. *Digital Spectral Analysis : with Applications.* Prentice Hall, 1987.

17. D Q Mayne and F Firoozan. Linear identification of arma processes. *Automatica*, 10(4):461–466, 1982.

18. A Papoulis. *Probability, Random Variables and Stochastic Processes.* McGraw Hill, 1984.

19. M B Priestley. *Spectral Analysis and Time Series. Volume 1 : Univariate Series.* Academic Press, 1981.

20. M B Priestley. *Spectral Analysis and Time Series. Volume 2 : Multivariate Series, Prediction and Control.* Academic Press, 1981.

21. J G Proakis and D G Manolakis. *Introduction to Digital Signal Processing.* Macmillan, 1987.

22. T Subba Rao, editor. *Developments in Time Series Analysis.* Chapman Hall, 1983. In honour of Maurice B. Priestley.

23. S E F Rofe. *Signal Processing Techniques for Plant Fault Detection.* PhD thesis, University of Queensland, 1996.

24. H Tong. *Non-linear Time Series. A Dynamical Systems Approach.* Oxford Science Publications, 1990.

25. Ronald J Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1:270–280, 1989.

Signal Processing Methods for Gearbox Fault Detection

Simon Rofe

AUSTRALIA

1.DEFENCE ORGANISATION

a.S&T Program

Chief Defence Scientist
FAS Science Policy
AS Science Industry and External Relations
AS Science Corporate Management
} shared copy

Counsellor Defence Science, London (Doc Data Sheet)
Counsellor Defence Science, Washington (Doc Data Sheet)
Scientific Adviser to MRDC Thailand (Doc Data Sheet)
Director General Scientific Advisers and Trials/Scientific Adviser Policy and Command (shared copy)
Navy Scientific Adviser (3 copies Doc Data Sheet and one copy of the distribution list)
Scientific Adviser - Army (Doc Data Sheet and distribution list only)
Air Force Scientific Adviser
Director Trials

Aeronautical and Maritime Research Laboratory
Director

Electronics and Surveillance Research Laboratory
Director

Chief of Airframes and Engines Division
Research Leader Propulsion
S.E.F. Rofe
A. Wong
B. Rebbechi
D. Forrester
D. Blunt
M. Burchill

DSTO Library
Library Fishermens Bend
Library Maribyrnong
Library DSTOS (2 copies)
Australian Archives

Library, MOD, Pyrmont (Doc Data sheet)

b. Forces Executive
Director General Force Development (Sea), (Doc Data Sheet)
Director General Force Development (Land), (Doc Data Sheet)
Director General Force Development (Air)

c. Army
ABCA Office, G-1-34, Russell Offices, Canberra (4 copies)

d. Air Force
DTA-LSA
ARDU

e. Navy
Peter O'Neil - NALO

f. S&I Program
Defence Intelligence Organisation
Library, Defence Signals Directorate (Doc Data Sheet only)

g. Acquisition and Logistics Program
No compulsory distribution

h. B&M Program (libraries)
OIC TRS, Defence Central Library
Officer in Charge, Document Exchange Centre (DEC), 1 copy
*US Defence Technical Information Centre, 2 copies
*UK Defence Research Information Center, 2 copies
*Canada Defence Scientific Information Service, 1 copy
*NZ Defence Information Centre, 1 copy
National Library of Australia, 1 copy

2.UNIVERSITIES AND COLLEGES

Australian Defence Force Academy
Library
Head of Aerospace and Mechanical Engineering
Deakin University, Serials Section (M list), Deakin University Library
Senior Librarian, Hargrave Library, Monash University
Librarian, Flinders University

3.OTHER ORGANISATIONS

NASA (Canberra)
AGPS

OUTSIDE AUSTRALIA

4.ABSTRACTING AND INFORMATION ORGANISATIONS
INSPEC: Acquisitions Section Institution of Electrical Engineers
Library, Chemical Abstracts Reference Service

Engineering Societies Library, US
Materials Information, Cambridge Scientific Abstracts, US
Documents Librarian, The Center for Research Libraries, US

5. **INFORMATION EXCHANGE AGREEMENT PARTNERS**
Acquisitions Unit, Science Reference and Information Service, UK
Library - Exchange Desk, National Institute of Standards and Technology, US
National Aerospace Laboratory, Japan
National Aerospace Laboratory, Netherlands

6. **ADDITIONAL DISTRIBUTION**
Chadwick - Helmuth Company Inc
4601 N. Arden Drive
El Monte, CA 91731-1299

SPARES (10 copies)

Total number of copies: 65

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE

Signal Processing Methods for Gearbox Fault Detection

3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document	(U)
Title	(U)
Abstract	(U)

4. AUTHOR(S)

Simon Rofe

5. CORPORATE AUTHOR

Aeronautical and Maritime Research Laboratory
PO Box 4331
Melbourne Vic 3001

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-0476	AR-010-104	Technical Report	February 1997

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
M1/9/271	DST 95/096	DSTO	39	25

13. DOWNGRADING/DELIMITING INSTRUCTIONS

14. RELEASE AUTHORITY

Chief, Airframes and Engines Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE CENTRE, DIS NETWORK OFFICE, DEPT OF DEFENCE, CAMPBELL PARK OFFICES, CANBERRA ACT 2600

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CASUAL ANNOUNCEMENT Yes

18. DEFTEST DESCRIPTORS

Helicopters, Gearboxes, Signal Processing, Fault Detection, Simulation, Vibration Analysis

19. ABSTRACT

Methods of accounting for load variation in vibration signals from helicopter transmission systems are presented. These methods are based on autoregreassive moving-average (ARMA) models, and several ARMA parameter estimation schemes are presented. Simulations of load variation are carried out, and a prediction error filter, based on the ARMA models, is used to generate a residual signal. Fault indices extracted from the residential signal are used to indicate the presence or absence of a fault. The results of the simulations suggest that this method of fault detection is able to detect both general and local fault conditions.